ABSTRACT OF THE DISCLOSURE

A technique for generating cross-references among categories in a knowledge base is disclosed. A plurality of themes are extracted from a corpus of documents. A theme identifies subject matter contained in a corresponding

5   document. A plurality of scores are generated such that each score identifies a relative theme strength among theme pairs of the themes extracted from the documents. In general, a theme strength reflects the amount of subject matter contained in a document for a corresponding theme relative to other themes in the document. Thereafter, the most related theme pairs are selected as indicated

10   by the scores. Category pairs of the knowledge base are then selected by mapping the themes of the selected theme pairs to corresponding categories of the knowledge base. A cross-reference between categories of the category pairs in the knowledge base are generated so as to identify an association between the category pairs.